



Vers une exo-mémoire personnelle supportée par un seul fichier texte

Pierre Deransart

► To cite this version:

Pierre Deransart. Vers une exo-mémoire personnelle supportée par un seul fichier texte. 2010. inria-00515437

HAL Id: inria-00515437

<https://hal.inria.fr/inria-00515437>

Preprint submitted on 6 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vers une exo-mémoire personnelle supportée par un seul fichier texte

Pierre Deransart

INRIA Rocquencourt, BP 105, 78153 Le Chesnay Cedex, France
`Pierre.Deransart@inria.fr`

Résumé Cet article ¹ présente une expérimentation avec une exo-mémoire textuelle continuellement mise à jour, dont le rôle est d'assister la mémoire naturelle d'un sujet. Il montre comment une théorie des traces pourraient être utilisés pour améliorer le dispositif. Les principales caractéristiques d'une telle mémoire sont qu'elle est exclusivement gérée par le sujet, de taille limitée, dotée de plasticité et de persistance dans la qualité du rappel à long terme.

1 Introduction

Cet article est un essai pour définir ce que pourrait être un artéfact numérique susceptible de servir d'exo-mémoire personnelle ², avec l'objectif non pas de remplacer notre mémoire naturelle mais simplement de la renforcer grâce à un outil. On ne cherche pas ici à pallier des déficits avérés ni traiter des pathologies liées au vieillissement, même si cette réflexion peut conduire à concevoir des artéfacts adaptés. On suppose au contraire que l'utilisateur n'a pas de déficit avéré, mais qu'il souhaite améliorer ses capacités pratiques de vie. Celles-ci nécessitent à tout moment de pouvoir disposer d'informations précises (sur une personne, un lieu, un événement passé, un dossier en cours, une affaire ancienne, ou une curiosité personnelle ...) et de savoir compléter les bribes d'informations dont sa mémoire n'a gardé que quelques traces trop insuffisantes pour être utilisables en pratique.

Un exemple simple et courant est un agenda nominal (papier ou électronique) structuré par l'ordre alphabétique des noms propres auxquels sont associées diverses informations comme numéro de téléphone, adresse, et autres détails éventuels. Un autre support courant est le cahier de notes de travail ou le journal personnel structuré par le temps et comportant une suite de paragraphes décrivant quelques événements du jour. Tant que le support est de type cahier

1. 30 août 2010 Une version légèrement réduite de cet article est publiée dans les actes du workshop PSD 2010 <http://ekaw2010.inesc-id.pt/psd.html>

2. ou mémoire externe. Un tel artéfact rentre dans la catégorie des HDM (Human Digital Memories) et a des liens avec les domaines PIA (Personal Information Archive), ou PIM (Personal Information Management).

physique, la récupération d’une information devient de plus en plus difficile car le volume total et l’ordre structurel imposé font que la récupération d’une information nécessite un recours à sa propre mémoire. Ainsi retrouver un numéro de téléphone dans un tel agenda peut s’avérer impossible ou très laborieux si on a oublié le nom de la personne à qui on veut téléphoner. Un cahier de notes peut devenir inutile car l’évolution des informations liées à des événements très liés entre eux mais très séparés dans le support (suite de compte-rendus de réunions échelonnées dans le temps sur un sujet donné) peut devenir pratiquement impossible.

Aujourd’hui existe une multitude de supports numériques qui offrent outre les fonctions primordiales évoquées ci-dessus, des services extrêmement variés comme par exemple l’agenda partagé collectif ou la saisie de notes au stylo, immédiatement numérisées et classées. Par ailleurs la croissance exponentielle, au-delà de la simple loi de Moore³ des moyens de stockage numériques, amène à penser que la quasi totalité des événements d’une vie puisse être stockée sur un seul disque dur. Dans [2] les auteurs estiment à environ 16 Go par an les besoins de stockage de tous les éléments de vie quotidiens d’une personne (incluant tous éléments de contexte de vie, courriels, sons, images, vidéos et musiques). Ceci, à l’échelle d’une vie, donne actuellement la possibilité de garder sur un support numérique à usage privé une trace très précise de l’ensemble des activités et interactions d’un seul individu. Même si l’accélération des moyens de communications et d’échanges laisse à penser que ces capacités peuvent être sous-estimées, il devient aujourd’hui possible d’envisager de conserver l’intégralité de ses “souvenirs” sur un support numérique individualisé avec des possibilités remarquables de pouvoir naviguer dans cet océan de données.

Il faut tout d’abord remarquer que le fait d’externaliser sa propre mémoire dans le but plus ou moins explicite de ne pas l’assister ou simplement de la conserver, occupe une partie importante de l’activité humaine. Prendre des notes, les classer, utiliser un agenda, constituer une bibliothèque, organiser son environnement participent à un tel but. Les personnes âgées vivent dans un environnement rempli d’objets souvenirs⁴ qui contribuent à leur qualité de vie, c’est à dire qu’ils les aident en partie à conserver une partie de leur mémoire. On voit ainsi que la simple automatisation d’une bibliothèque relève de l’utilisation d’un artefact numérique à des fins d’amélioration d’une mémoire individuelle ou collective. Si l’idée d’auxiliaire de mémoire peut englober en fait de vastes systèmes sociaux, nous nous limitons ici à ce qui est usuellement appelé “mémoire personnelle” plus proche d’un agenda perfectionné susceptible d’accompagner la personne à chaque moment de sa vie.

3. Une version de la loi dite “de Moore” stipule que la puissance des processeurs double tous les 2 ans depuis 1969. Cette croissance exponentielle est trois à quatre fois plus forte pour les moyens de stockage numériques [1].

4. Par “objets souvenirs” on entend ici une collection d’objets éventuellement pourvus de charges émotionnelles liées à des événements significatifs de vie et qui aident à en garder la mémoire. Ceci est remarquablement décrit dans le roman de D.Coulin [3].

Cet article comporte quatre sections principales. Dans la première (section 2) nous caractérisons ce que nous entendons par exo-mémoire ou auxiliaire de mémoire personnel. La seconde (section 3) décrit une expérimentation avec une exo-mémoire représentée par un fichier textuel sous éditeur de texte. La troisième (section 4) donne quelques fondements théoriques possibles utilisant la notion de trace, et la dernière (section 5) discute quelques points caractéristiques essentiels de ce type d'exo-mémoire ainsi que ses rapports avec d'autres approches.

2 Exo-mémoire personnelle

Dans cette section on tente de cerner les caractéristiques essentielles de ce que nous appelons ici exo-mémoire.

En 1945 Vannevar Bush a écrit un article intitulé “As We May Think” dans lequel il jetait les bases de Memex [4] : “un appareil dans lequel un individu peut stocker tous ses livres, musiques et autres éléments de communication, et mécanisé de telle sorte qu'il peut être consulté très rapidement et de manière très flexible”.

Aujourd'hui cet article revêt encore un aspect prophétique. Un sentiment domine qui donne l'impression que l'apparition du WWW donne en particulier un accès quasi illimité à une sorte de mémoire universelle qui contiendrait l'ensemble des connaissances anciennes et actualisées. Le développement des services et en particulier la possibilité d'accéder à des informations décentralisées quelle que soit leur localisation donne le sentiment qu'il suffit de stocker des liens pour que notre mémoire, essentiellement constituée de connaissances acquises (par nous ou par d'autres - forme d'éducation collective) soit “automatiquement” approvisionnée, nous dotant d'une sorte de cerveau amplifié.

Cette vue serait tout à fait correcte si on limitait la fonction de la mémoire au stockage de connaissances et à la capacité de les retrouver rapidement. On a bien compris aujourd'hui que, la mémoire humaine ne peut être vue simplement comme des mécanisme d'accès à des informations accumulées sans arrêt ni limite. Malgré toute sa richesse combinatoire, aucun cerveau humain n'aurait une capacité de stockage suffisante. La mémoire relève de capacités particulières de stockage à plus ou moins long terme, d'abstraction, de rappel et surtout d'oubli.

Sans rentrer dans les détails des processus biologiques mis en jeu pour la mémoire humaine, il est utile d'en apprécier les propriétés fondamentales pour pouvoir parler d'exo-mémoire.

D'une manière tout à fait classique, comme le résume G.Chapouthier [5] p.64 et suivantes, on peut caractériser la mémoire humaine selon trois axes : sensoriel, temporel et abstrait. Sur l'*axe sensoriel* les sensations tactiles, auditives, visuelles, olfactives, etc. L'*axe temporel* se réfère à la rémanence du souvenir : brève est la mémoire de travail (au plus quelques minutes), dite aussi épisodique ou mémoire transitoire ; plus durable (de quelques heures à plusieurs années) la mémoire dite de référence. Celle-ci correspond aux acquis stables. Enfin sur l'*axe abstrait* s'opposent la mémoire procédurale ou implicite (les habitudes acquises)

et la mémoire déclarative ou explicite (les significations). Dans la première le rappel est spontané et immédiat (mouvement travaillé du geste sportif par exemple mais dont la réalisation correcte se fait selon des circuits neuronaux rapides), alors que dans la seconde un appel à la réflexion est nécessaire. C'est celle qui nécessite en particulier le recours à des formes de raisonnement.

Ainsi sommairement classées, chaque mémoire a son mode d'utilisation ou de *rappel*. Par exemple pour la mémoire implicite, le rappel est inconscient et "automatique" -c'est un "savoir comment"-, alors que pour la mémoire explicite, le rappel est conscient et relève de l'application de règles (lesquelles auraient une fonction plus sémantique) -c'est un "savoir que"-. Ces mécanismes de rappel, plus ou moins rapides sont associés à un phénomène d'oubli qui opère un tri dans les deux sens sur ce qui doit être "enregistré" ou non, et ce qui doit être récupéré ou pris en compte.

Notre but ici n'est pas de chercher à réaliser un modèle de la mémoire vivante évoquée ci-dessus, mais essentiellement de rechercher des formes d'extension "mécaniques" non vivantes, qui permettraient de la renforcer. Une manière d'aborder cette question est de ne considérer de la mémoire vivante que ce qui peut être externalisé, et éventuellement traité numériquement.

On va donc s'intéresser à une forme de mémoire correspondant à des informations digitalisées (axe sensoriel), durables (axe temporel), et essentiellement conscientes (axe abstrait).

- Axe sensoriel : Toute sensation qui peut être numérisée peut être stockée sur un support numérique. Ces informations peuvent s'accumuler automatiquement sur un ordinateur ou tout autre support numérique plus spécifique, ou simplement provenir du butinage du sujet. On peut aussi considérer que l'information reste sur un support diffus (par exemple vidéo sur YouTube⁵). Ces questions font l'objet de nombreuses recherches (voir en particulier la série de colloques [6]). Le périmètre support de l'exo-mémoire peut s'étendre d'un simple fichier, au WWW tout entier en passant par un ordinateur et son système d'exploitation.
- Axe temporel : La mémoire de travail n'a pas à être distinguée ici de la mémoire à long terme. Tout événement mémorisé peut avoir une marque de temps. La question de savoir s'il s'agit d'un événement propre à la mémoire de travail ou de référence est a priori une simple question d'ancienneté de l'événement. C'est aussi une question d'interface avec le sujet. Il y a peu de chance, dans l'état actuel des techniques, que le sujet puisse utiliser son exo-mémoire assez rapidement ; par exemple, au milieu d'un discours, pour retrouver un mot qui lui échappe ! L'exo-mémoire agira donc surtout comme mémoire de référence.
- Axe abstrait : il n'y a pas non plus à distinguer les traitements rapides inconscients des traitements conscients plus lents ; cette distinction est plus physiologique que calculatoire. En effet la distinction entre une mémoire dont le rappel est automatique et donc reposerait éventuellement sur un

5. [http ://www.youtube.com/](http://www.youtube.com/)

algorithmes, et une mémoire dont le rappel est explicite et donc reposerait sur des règles et leur utilisation dans un “raisonnement”, correspond seulement à une représentation différente des calculs à effectuer. La distinction ici est juste une question de codage. Ceci ne veut pas dire que certains traitements opérés au sein de l'exo-mémoire pour effectuer des rappels ne puissent être plus ou moins efficaces selon les algorithmes utilisés.

Il faut également ajouter quelques particularités fondamentales de ce que nous estimons pouvoir être une exo-mémoire.

- Privé vs. Public : une exo-mémoire est un artéfact essentiellement privé. Son isolement du monde extérieur et son accès réservé au seul sujet sont des atouts fondamentaux pour qu'elle puisse fonctionner. Une telle mémoire n'est en effet utile que si le sujet peut y mettre ce qu'il veut. Par exemple ses codes secrets de plus en plus nombreux et dont on demande qu'ils soient si possibles aussi différents les uns des autres. Il doit pouvoir y mettre, pour des raisons mnémoniques qui lui sont propres, ce qu'il pense vraiment - afin que le rappel ne soit pas contraint - par exemple pouvoir associer au nom enregistré (mais dont on suppose qu'il puisse l'oublier) d'un véritable ami, des expressions telles que “dent gâtée” ou “puant de la bouche”, sans craindre que ceci puisse être mal interprété par une tierce personne⁶.

Cet aspect de l'exo-mémoire lui confère actuellement quelques limitations expérimentales. Ceci est bien mis en évidence dans [7] où les aspects légaux et sociaux doivent être pris en considération pour pouvoir mener des expériences autres que personnelles. Par ailleurs l'inviolabilité du support doit pouvoir être assurée. Dans l'état actuel des techniques, ceci est impossible⁷. Une manière de l'approcher est, en particulier, de limiter l'exo-mémoire à un fichier de taille réduite qui peut être crypté sur un ordinateur (ou autre artéfact) personnel.

Il faut noter enfin que le fait d'être privée confère à l'exo-mémoire sa souplesse et facilite la résolution de problèmes formels qu'elle peut poser. En effet ce n'est pas la même chose de devoir concevoir des ontologies⁸ susceptibles d'un usage universel et une ontologie privée. À partir du moment où l'usage de l'exo-mémoire est strictement réservé à une personne, même si sur un plan théorique les problèmes sémantiques éventuels gardent la même complexité (par exemple reconnaissance d'une expression), sur un plan pratique, des dimensions et donc des ordres de grandeur de difficulté peuvent disparaître (par exemple la prise en compte des aspects multiculturels dans la notation des dates est a priori inutile dans une exo-mémoire). Il est tout à fait normal que le contenu d'une exo-mémoire soit incompréhensible pour

6. Une telle exo-mémoire ne prend vraiment sens que si son sujet, comme pour sa mémoire biologique, peut disparaître avec elle, sauf demande explicite, par exemple dans son testament. Elle devrait également être considérée comme inviolable, même par la force publique.

7. Ce point mériterait un ample développement, mais il est orthogonal à cet article.

8. Voir discussion.

tout autre personne que son sujet, sa mise à jour s'apparentant en fait à de la prise de notes personnelles.

- Manipulations automatiques vs humaines : l'exo-mémoire ne peut se limiter à accumuler des données fussent-elles obtenues à partir de toutes sortes de capteurs fixés sur le sujet. Dans un premier temps au moins (nous reviendrons sur cet aspect ultérieurement), elle ne peut qu'être contrôlée par le sujet. Seul celui-ci peut sélectionner les événements qu'il juge utiles, les introduire et les annoter de manière à pouvoir les retrouver quand il les aura oubliés. Il ne peut donc être question de confier a priori cette tâche à un automate, car seul le sujet, en accord avec sa mémoire biologique, peut accomplir les choix nécessaires sur l'instant. Peut-être cela ne serait-il plus possible quelques secondes ou minutes plus tard.

Il n'est pas non plus possible de demander au sujet de sélectionner dans une grande collection d'enregistrements réalisés automatiquement les parties significatives et les annoter avec des catégories prédéfinies (si personnelles et adaptées qu'elles puissent être). Ce type de travail demanderait trop de temps. En effet un problème majeur est que l'exo-mémoire doit rester une aide et non un devoir fastidieux ou simplement impossible à mener parce que le sujet n'en a simplement pas le temps. Même s'il peut être intéressant de rentrer le plus de faits possibles, on ne cherchera pas à être exhaustif, car cela est de toute manière impossible.

3 Expérimentation

Nous relatons ici une expérimentation avec une exo-mémoire dont l'organisation se fait à travers un fichier texte utilisé et maintenu pendant plusieurs années avec pour seul éditeur "emacs", l'usage de la seule fonction "search" comme fonction de rappel, et pour seul sujet l'auteur de cet article. La description est un peu simplifiée, mais reflète l'essentiel de la structure de l'artéfact et des comportements du sujet.

L'unité d'information de base est la ligne (de longueur quelconque), appelée *m-ligne*, *m* pour "mémo". Des m-lignes peuvent être ajoutées les unes après les autres ou insérées au milieu d'autres; une m-ligne peut-être simplement modifiée. Le principe est qu'une m-ligne a un seul thème principal. Une m-ligne est constitué de suites de mots ou signes séparés par une virgule ou un séparateur quelconque. Les m-lignes sont séparées les unes des autres par le simple fait de commencer en début de ligne sans aucun interligne. Chaque m-ligne est sensée correspondre à ou stimuler un souvenir jugé utile ou important.

Le texte est écrit sans formatage particulier et toutes sortes d'informations peuvent y être insérées. Si l'on veut introduire une information qui n'est pas textuelle (image, son, document volumineux) on ne rentre que les métadonnées textuelles (cf. plus bas) et une indication de l'endroit où l'on peut la trouver (espace ordinateur personnel, espace physique domestique ou espace WWW). L'éditeur de texte sert ici également de moteur de recherche. A tout instant

il est possible de faire une recherche pour vérifier certaines cohérences entre différentes m-lignes. Il est important de noter que le moteur de recherche n'est pas utilisé pour simplement retrouver des informations mais pour retrouver un souvenir. Nous l'appellerons donc "moteur de rappel".

Le fichier complet sera appelé *m-book* pour *mémo-book*. Un certain nombre de précautions doivent cependant être prises pour que le moteur de rappel puisse fonctionner de manière efficace.

- **Écriture normalisées de certains éléments.** Afin de pouvoir être retrouvés sans difficulté, des passages de textes doivent être écrits sans faute d'orthographe⁹. Les noms propres doivent être correctement écrits, les numéros de téléphone doivent avoir la même forme, ..., bref une certaine normalisation du texte est nécessaire. Cet effort de normalisation est nécessaire pour certaines unités qui vont jouer le rôle de mots clef. Par exemple il faut pouvoir faire une recherche inverse sur un numéro de téléphone. Ces "normes" cependant peuvent rester personnelles. La seule raison d'une telle normalisation est de permettre au sujet de mener une recherche inverse. Il doit utiliser les normes qu'il connaît précisément ou qu'il est capable de mémoriser sur le long terme.
- **Métadonnée.** On appelle ici métadonnée toute information qui va aider au rappel d'une m-ligne. Une grande liberté est laissée au niveau des métadonnées qui peuvent même être formulées dans un langage incompréhensible ou non structuré. Ces données peuvent être mises n'importe où dans une m-ligne et sont en général redondantes. Par exemple, le rappel du nom d'une personne que l'on recherche peut se faire d'abord avec le nom ou d'abord avec le prénom. On ne peut en général prévoir ce qui viendra en premier, surtout si on n'a plus eu aucune relation avec cette personne depuis plusieurs années. Si une m-ligne traite de **Pierre Deransart** on pourra mettre en métadonnée :
Pierre deransart pierre ou **deransart pierre deransart**.
 De manière plus générale, on peut mettre dans les métadonnées des informations contextuelles qui peuvent aider à retrouver cette personne à l'aide d'autres attributs. Ainsi par exemple, on pourra entrer :
Pierre deransart pierre imberbe pierre yeux clairs yeux fait sport pierre sport pierre iclp pierre programmation logique deransart lp
 Le principe est de mettre le plus d'attributs possibles en essayant de se représenter comment il sera possible de retrouver ce **pierre** lorsque j'aurais tout oublié de lui après plusieurs années.
- **Constitution de m-paragraphe et m-pages.** Au fur et à mesure que la taille d'une m-ligne augmente il devient plus difficile en pratique de lire les informations qu'elle contient. Il peut alors être plus agréable de la scinder en plusieurs m-lignes. Par exemple, après quelques années, on a pu accumuler plusieurs types d'informations concernant une per-

9. Au moins avec une orthographe constante pour le sujet. Avec un moteur de rappel plus sophistiqué, on peut limiter cette exigence.

sonne, comme localisation(s) géographique(s), moyens de contact (email, téléphones, fax, ...), composition de la famille, intérêts, rencontres (sorties marquantes, disputes, ...), profession, actions remarquables, photos, publications,..., dont on a estimé qu'il pouvait être utile de conserver la mémoire sans pour autant créer un support particulier d'archivage. On obtient alors une meilleure lisibilité en coupant la m-ligne existante en plusieurs m-lignes qui retrouvent une meilleure homogénéité et par là même une meilleure lisibilité. Ces m-lignes ne sont séparées par aucun espace. Ce nouvel ensemble de m-lignes sans interligne constitue un m-paragraphe thématique dont l'évolution dans le temps se traduit par enrichissement de taille et nombres des m-lignes. De temps en temps, si le thème s'amplifie, de nouvelles m-lignes peuvent être créés ou coupées.

Voici un exemple peu compromettant de m-paragraphe court avec m-lignes et métadonnées concernant la gestion de la machine à café de l'équipe "contraintes" de l'INRIA où le sujet travaille.

```
CAFE inria cafe machine cafe inria cafe depannage cafe
contraintes cafe projet cafe commande cafe
-Societe D8, 7-8 rue Leon Geffroy, 94408 Vitry Cedex
-Client 30700
-commande: 01 47 18 38 40, par 200 pour 70 euros, 3 cat:
fort (brun), moyen (vert) et faible
-depannage: 01 47 18 38 30 7h30-17h (9h samedi)
-matricule appareil: 034726
-commande 9/4/10: 35 euros (cafe 100 doses, gobelets, spatules
et sucres, sinon 32 euros) livraison a l'occasion (a partir du
lundi 12)
```

La première m-ligne fixe le thème principal. L'orthographe et la syntaxe peuvent être simplifiés selon des règles propres au sujet (absence d'accents), les m-lignes sont ici très courtes et traitent soit de données de la machine, soit d'incidents ou de commandes. Les métadonnées de la première m-ligne ont été introduites pour améliorer des recherches laborieuses antérieures de ce paragraphe et correspondent à différentes associations possibles contextuelles du thème : "café inria" (lieu de travail) ou "café contraintes" (équipe projet) ou "café projet" ou "café commande".

Comme il n'y a pas de stratégie de croissance de l'exo-mémoire, plusieurs m-paragraphe thématiques sur des sujets proches peuvent avoir été créés. Il est alors possible de les regrouper dans le fichier de manière à appréhender plus facilement une thématique plus large. Ici le support physique atteint ses limites tel qu'il est utilisé car bien sûr plusieurs regroupements sont possibles et ce dispositif n'a pas pour but de permettre toutes sortes de regroupements. Ce point sera discuté ultérieurement. Ce regroupement est donc purement occasionnel, mais il peut s'obtenir également en ajoutant des métadonnées permettant d'identifier et faire défiler les m-paragraphe correspondant à un thème d'intérêt particulier. Un tel re-

groupement s'appellera *m-page* et bien sûr il y a toutes sortes de m-pages possibles ; c'est le sujet qui décide s'il y a lieu d'en créer une ou non, ou de réorganiser les m-pages.

- **Traitement des données caduques.** Certaines données deviennent caduques (par exemple changement d'adresse d'un ami) ; certaines sont simplement incertaines (nom mal entendu au téléphone ou âge d'une personne non connu avec certitude). Comme il s'agit non de traitement formel, mais de traitement des données par le sujet en état de rappel, on se contente d'utiliser les symboles : ? pour l'incertitude, et % pour la donnée devenue caduque. L'opérateur “%” est préfixe et s'applique à toute la phrase. L'opérateur “?” s'accroche à un mot avant ou après et selon le contexte son influence est limitée au mot ou à la phrase.

Les données caduques sont donc conservées et parfois déplacées à la fin d'un m-paragraphe si elles gênent trop la lisibilité. Les garder peut présenter un certain intérêt pour le rappel. Une question néanmoins concerne l'effacement complet de données (par exemple éventuellement le changement d'URL d'une page web). Un tel effacement reste rare, car peu dérangeants en général pour le rappel ou la lisibilité.

L'évolution du m-book se fait donc à l'aide des opérations suivantes. Pour les m-lignes : création (*create_m-line*), insertion de données ou métadonnées (*insert_m-line_content*), corrections d'ordre linguistique, marquage des données incertaines (*dead_content*) ou caduques (*uncertain_content*) ; pour les m-paragraphe : création par coupure ou regroupement de m-lignes (*create_m-paragraphe*), fusion (*fusion_m-paragraphe*) ou coupure (*split_m-paragraphe*) de m-paragraphe ; pour les m-pages : création par regroupement physique de m-paragraphe (*create_m-page*) ou par ajout de métadonnées dans les m-paragraphe qui le composent, regroupement physique de m-pages (*fusion_m-page*).

Noter que l'on pourrait multiplier les niveaux de structures indéfiniment (m-chapitres, ... etc.). Le support considéré ne permet guère d'aller au-delà de deux niveaux et ceci de manière insatisfaisante puisque la création d'une m-page peut en désorganiser une autre. En pratique cette possibilité n'est pas vraiment utile car, poussée à l'extrême, elle aboutirait à imposer une structure globale au m-book, sans certitude que cette structure garde la même cohérence dans la durée puisque le sujet se construit et évolue en même temps que sa mémoire.

Nous expérimentons depuis 6 ans sur un fichier personnel qui constitue une exo-mémoire en pratique très utile, rapide, efficace et surtout persistante dans le sens où le rappel ne se dégrade pas avec le temps. La stratégie qui permet d'éviter cette dégradation et d'améliorer continuellement l'efficacité du rappel consiste à ajouter systématiquement des métadonnées chaque fois qu'une m-ligne ou un m-paragraphe n'est pas rappelé immédiatement avec quelques mots clefs ou une seule expression, mais nécessite une recherche plus ou moins longue. Une telle métadonnée ne peut être inférée automatiquement dans la mesure où les mots ou expressions que l'on va vouloir ajouter (comme une sorte de

raccourci mnémonique) sont des termes jaillis de la mémoire du sujet et sont souvent imprévisibles.

Avec un usage systématique, la croissance du m-book est de l'ordre de 200 KB par an. Cette croissance est linéaire et non exponentielle dans la mesure où le choix des informations à y garder est exclusivement manuel et que la taille de la saisie est en moyenne proportionnelle au temps passé à introduire de nouvelles informations ou à les mettre à jour. La taille est par ailleurs limitée par le fait que, même s'il est possible d'introduire des portions de texte par copier/coller, en général seuls des pointeurs ou références sur des données plus importantes sont entrés dans le m-book. Il est en effet essentiel que les m-lignes soient le plus court possible dans un m-paragraphe. On voit donc qu'à l'échelle d'une vie la taille du m-book (de l'ordre de quelques dizaines de méga octets) ne peut être un obstacle réel à l'efficacité de nombreuses applications possibles, en particulier de moteurs de rappel.

La rapidité de croissance serait tout autre si le même m-book était construit de manière automatique, par exemple à partir d'une ontologie personnelle rendant compte d'intérêts personnels potentiels. Ceci ne ferait que refléter la croissance exponentielle des connaissances globales, mais correspond à un tout autre problème. Il resterait de toute manière à opérer un tri afin de ne retenir que les informations jugées suffisamment fiables et significatives sur un plan personnel. Ce que seul le sujet peut réaliser.

4 Modélisation

Nous présentons brièvement une approche théorique possible fondée sur la notion de trace telle que présentée dans [8,9,10] et issue du génie logiciel.

L'idée principale est que le m-book tel que décrit ci-dessus n'est qu'une représentation possible de l'état d'un système (la mémoire) qui a été modélisé par une série d'actions qui ne sont connues que par la suite des événements (ou *trace*) qui ont permis de l'obtenir. Cette trace peut être formalisée par une *trace effective*¹⁰. L'état de la mémoire à un instant donné, dont le m-book n'est qu'une présentation visuelle possible, est dit *état virtuel* de la mémoire. A l'instant zéro, l'état virtuel *initial*, comme la mémoire, est supposé vide. L'état virtuel *courant* (au-delà de l'état initial) ne peut être complètement connu qu'à partir d'une trace effective.

Une trace effective est une suite d'événements de trace de la forme $w_t : (t, a_t)$, $t \geq 1$ où t est le chrono (un entier incrémenté à chaque événement) et a_t dénote une suite finie de valeurs d'attributs. a_t est l'état effectif courant. Chaque état a_t contient au plus n attributs dont le nombre dépend exclusivement du type d'action l'ayant produit. Une trace effective finie de t ($t > 0$) événements est dénotée $T_t^w = \langle s_0, \overline{w}_t \rangle$, où s_0 est l'état virtuel initial commun aux deux traces et \overline{w}_t représente la suite $w_1, \dots, w_i, \dots, w_t$.

10. Trace effective intégrale contigue dans [8].

Les attributs possibles de la trace “exo-mémoire” sont : un type d’action (*create_m-line*, *insert_m-line_content*, *dead_content*, *uncertain_content*, *create_m-paragraphe*, *fusion_m-paragraphe*, *split_m-paragraphe*, *create_m-page*, *fusion_m-page*) et la description des modifications du m-book (par exemple, pour un ajout de texte dans une m-ligne : portions de texte ajouté, m-ligne concernée et position dans la m-ligne). La position dans la m-ligne est essentielle car le contexte d’insertion contribue également au sens que l’on veut attribuer au texte.

La sémantique associée à une telle trace est une sémantique de reconstruction ou *sémantique interprétative* (SI) qui à partir d’une telle trace permet d’obtenir l’état virtuel courant. On voit donc qu’à toute trace finie de taille t , $T_t^w = \langle s_0, \overline{w_t} \rangle$, on peut associer un état virtuel s_t . La suite $s_0, s_1, s_2, \dots, s_t$ constitue la suite des états virtuels atteints par la mémoire au cours de sa “vie”. Si on note a_t la dernière action effectuée pour atteindre l’état s_t , alors la suite $T_t^v = \langle s_0, \overline{(a_t, s_t)} \rangle$ peut également être vue comme une trace. Elle est dite trace virtuelle car le deuxième attribut est l’état virtuel.

La SI est ici donnée avec la trace. C’est elle qui permet d’interpréter¹¹ la trace en la représentant sous la forme d’un m-book. Il est important de bien distinguer ce que peuvent être l’interprétation d’une trace qui est une structure de donnée abstraite (portions de texte ordonnées par un système de pointeurs) et les différentes représentations que l’on peut en faire. Ici on se limite à une représentation à deux dimensions sous forme d’un m-book de lignes textuelles, mais on pourrait imaginer des applications permettant des représentations plus sophistiquées incluant plusieurs dimensions possibles et offrant la possibilité de plusieurs m-paginations.

La sémantique associée à la trace virtuelle, le moteur qui produit la trace effective, est appelée sémantique observationnelle (SO). Selon [8], les deux sémantiques (SI et SO) sont équivalentes et que la SO existe toujours théoriquement. La figure 1, issue de [8], illustre cette équivalence.

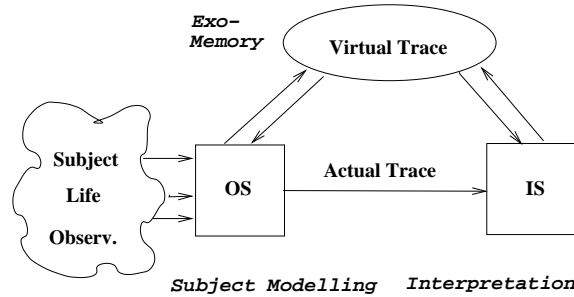


Figure 1. Modélisation

11. Pour cette raison, elle est appelée *sémantique interprétative* dans [8].

Dans ce schéma le “Subject Life Observation” est le monde en interaction avec le sujet qui lui-même modélise une trace virtuelle (l’exo-mémoire représentée ici sous forme d’un fichier) mais transcrit au jour le jour et de manière consciente une trace actuelle. Le fait qu’il existe une SO sur un plan formel ne signifie pas que l’on puisse remplacer le sujet par un automate, car décrire formellement cette SO reviendrait à modéliser au moins en partie aussi le monde dans lequel il vit. Cela ouvre cependant des perspectives sur des manières possibles de procéder afin, non de supprimer le sujet, mais de l’assister dans la collecte d’informations qu’il souhaite gérer dans son exo-mémoire ; par exemple, en utilisant des résultats d’une analyse sémantique d’un m-paragraphe, lui proposer d’introduire de nouvelles m-lignes potentiellement pertinentes.

Dans cette approche on peut considérer que toute m-ligne, tout m-paragraphe ou m-page, . . . , est une sous-trace. Le m-book peut alors être vu comme une base de traces au sens de J.-C. Marty et A. Mille [11,12] dont la trace effective décrite ici constitue une *trace première*. Cela ouvre la perspective de pouvoir utiliser des méthodes d’apprentissage dans un cadre théorique bien fondé afin d’aider le sujet à utiliser son exo-mémoire pour l’assister dans des découvertes de nouvelles connaissances en se basant sur sa propre expérience.

5 Discussion

On cherche à situer ce travail par rapport aux recherches sur la mémoire biologique, sur l’ingénierie des connaissances et sur les logiciels de gestion de notes.

Sur un plan neuro-biologique G. M. Edelman et G. Tononi [13] rappellent que la mémoire est essentiellement non représentationnelle¹² et que le cerveau est pourvu en fait de centaines voire des milliers de centres mnémoniques qui interagissent. Notre exo-mémoire rend possible de conserver de manière durable des bribes de souvenirs concernant a priori toutes sortes de centres mémoriels, mais elle a quelques limites. Ainsi sur ce que nous avons nommé l’axe sensoriel le mode de saisie des informations privilégie certains sens. Certaines sensations liées au contexte de réalisation d’un événement et pouvant contribuer au rappel d’un souvenir ne seront pas saisies faute de temps ou parce que la relation est justement inconsciente ; par exemple, comme il est noté dans les expérimentations de [2], l’influence de la couleur d’un document ou du temps qu’il faisait au moment de l’enregistrement d’un fait. Ce type d’information sera rarement noté avec ce type d’exo-mémoire textuelle, essentiellement du fait que l’acte d’écriture sur un support physique est en soi un acte conscient.

D’une certaine manière l’exo-mémoire privilégie l’axe abstrait. Il est cependant probable que l’acte d’écriture d’un fait dans l’exo-mémoire serve aussi par lui-même à sa mémorisation cérébrale et intervienne pour une part sans doute

12. Cela signifie en particulier que l’on ne mémorise pas tous les détails d’une scène, mais seulement quelques éléments pour la reconstruire partiellement en fonction de besoins particuliers.

dans la mémorisation inconsciente. Mais il a aussi un aspect contraignant qui peut contribuer à limiter l'acte même de stockage de souvenirs.

Dans le domaine de l'assistance médicale auprès de patients avec des déficits de mémoire avérés [14] la saisie des "souvenirs", c'est-à-dire la sélection des faits intéressants, se fait avec l'aide du personnel soignant ou d'un proche. La problématique comme le résultat que l'on peut en espérer sont essentiellement différents.

Dans [7] les auteurs passent en revue plusieurs systèmes d'enregistrement continu ou volontaires de souvenirs ou documents et décrivent le système iRemember [15], testé comme aide mémoire personnel spécialisé sur des enregistrements audio de fragments de vie (conversations, conférences, ...). Certains systèmes sont plus particulièrement destinés à compenser un défaut particulier de la mémoire du sujet ; d'autres sont plus centrés sur la gestion d'archives personnelles ; mais la plupart enregistrent automatiquement des événements de vie en continu y compris du contexte (par exemple, dans iRemember des données de position (GPS) du sujet ou météorologiques (WWW sites) peuvent être associées à chaque événement enregistré). Ces approches sont différentes de celles présentées ici dans la mesure où aucun enregistrement ne se fait automatiquement dans l'exo-mémoire, en tous cas pas sans le contrôle immédiat du sujet.

Les travaux concernant l'ingénierie des connaissances sont plutôt orientés vers la mémoire sociale, c'est-à-dire vers le stockage et le partage de connaissances au sein de groupes d'individus aussi divers que la famille, une entreprise, un réseau social ou même l'humanité entière dans un cadre de mondialisation. Ainsi dans le "Handbook of Research on Emerging Rule-Based Languages and Technologies" [16] la plupart des travaux concernent la réalisation si possible automatique d'ontologies destinées à la gestion d'archives dans différents contextes sociaux. Ces approches incluent fréquemment des réseaux ou mémoires sémantique, telles qu'introduits par Quillian dès 1968 [17]. Il est important d'observer que de tels réseaux peuvent être construits sur la mémoire virtuelle représentée dans le support physique. Ils peuvent servir de base à de multiples applications susceptibles de faciliter l'accès et la gestion de l'exo-mémoire. Mais, afin de préserver la plasticité de l'exo-mémoire, de tels réseaux ne doivent être construits et utilisés que progressivement, et surtout ne peuvent être imposés a priori.

Le cas des logiciels destinés à gérer des notes, voire l'ensemble des supports d'activités du sujet, comme EVERNOTE, DEVONthink, CintaNotes, SOHO notes 8, Yojimbo, ShoveBox, ou wikidPad, entre autres, correspond à une situation intermédiaire où on vise à faciliter l'organisation non seulement de quelques notes relatives à la vie du sujet, mais dans une certaine mesure, de l'ensemble des documents qu'il doit manipuler. Ces logiciels imposent en général une structure qui, à un certain stade de développement, peut devenir contraignante. Nous ne pouvons ici discuter chaque système, mais notre approche a un avantage sérieux, celui de la simplicité des logiciels suffisants pour la rendre utile. En effet de nombreux systèmes proposés sont extrêmement sophistiqués et l'utilisateur peut devenir dépendant d'un fournisseur. En se basant sur un système dont les fonctions

peuvent se réduire à un “simple” traitement de texte, celma permet au sujet d’être pratiquement sûr de toujours pouvoir utiliser son exo-mémoire.

Dans notre approche, la mémorisation sur le support numérique comporte en fait deux parties : l’exo-mémoire du sujet et l’ensemble des documents stockés par le sujet et constituants des archives personnelles. L’exo-mémoire, réduite à un fichier unique, peut agir comme point de passage pour aider à retrouver des documents archivés, jouant le rôle d’un thésaurus pour une encyclopédie. Autant l’exo-mémoire doit avoir un statut exclusivement privé (propriété privée et accès exclusif du sujet) et son mode de manipulation limite de fait sa croissance ; autant le statut des archives est nécessairement différent en raison même de leur mode de croissance qui peut être à la fois partagé, automatisé et exponentiel. De ce fait son statut privé et relativement autonome sur le plan sémantique ne peut plus être garanti. On voit ainsi qu’il y a deux domaines distincts de recherche et que les travaux qui s’y réfèrent, tout en gardant certains liens, sont de nature différente.

Concernant l’exo-mémoire, le travail de A.Kiss et J.Quinqueton sur le système Uniscript [18] est l’un de ceux qui se rapprochent le plus de l’exo-mémoire. Dans ce système, les événements de trace ont la forme de “stances” situées dans un temps et un espace. Ce sont des fragments significatifs d’histoire personnelle, des fragments contextuels ou des résumés de tranches de vie, qui contiennent des aspects événementiels codables dans un langage et avec des données digitalisées. Ces stances sont organisées en une sorte de réseau sémantique avec une logique propre. Ce réseau sert de méthodes d’accès aux archives. Le système inclut donc l’ensemble de la problématique exo-mémoire et archives personnelles. Cependant l’implantation qui a été réalisée, sous forme d’une base généalogique de souvenirs personnels, est gérée comme une application WEB et orientée vers la réalisation d’archives familiales.

6 Conclusion

Nous avons montré comment une trace écrite, consciente, comportant autant de faits de vie que possibles jugés dignes d’être conservés, spontanément ou après mures réflexions, construite inlassablement, contrôlée et organisée par son sujet, pouvait constituer une exo-mémoire. Nous avons montré également comment un fichier textuel, très grossièrement structuré et manipulé à l’aide d’un éditeur de texte, pouvait en constituer une approximation utile et efficace, qui inclut quelques aspects de plasticité et d’oubli propres à la mémoire neuronale. Enfin, du fait que la suite des mises à jour peut être vue comme une trace première, cette approche permet le développement d’utilitaires susceptibles d’améliorer les performances de l’exo-mémoire grâce à des interfaces qui en facilitent l’usage. De multiples perfectionnements deviennent en effet possibles en s’appuyant sur des outils existants conjuguant, en particulier, des techniques de bases de données, fouille de données, interprétation abstraite et récupération d’informations. De telles applications se retrouvent en partie dans de nombreux logiciels de gestion de notes.

L'originalité de cette approche réside dans les caractéristiques essentielles de cette forme d'exo-mémoire : simplicité et efficacité, fonctionnalités mnémoniques et créatrices, plasticité et sentiment de satisfaction. Ce dernier point est particulièrement important. Dans la mesure où le sujet a très souvent le sentiment qu'il "retrouve la mémoire" grâce à cet outil, ou au moins qu'il ne la perd pas, son utilisation constitue un renforcement positif pour son utilisation ; à condition toutefois que ce sentiment ne soit pas contrebalancé par une trop grande difficulté d'utilisation, et que le temps perdu à saisir les données soit clairement contrebalancé par le temps gagné à les retrouver.

Le premier point (simplicité et efficacité) est également essentiel pour l'accessibilité, c'est-à-dire pour un usage par des non spécialistes du clavier, de la souris et d'un éditeur de texte. C'est sans doute une difficulté technique majeure à surmonter, mais on peut penser que des méthodes de saisie toujours volontaires mais polysensorielles peuvent être adaptées pour gagner en facilité et efficacité (saisie orale par exemple ou en écriture manuscrite). Ce point est également essentiel pour la pérennité à long terme de l'exo-mémoire en lui permettant de ne pas dépendre exclusivement d'un seul fournisseur.

Quant aux fonctionnalités mnémoniques et créatrices, elles sont assurées par l'acte volontaire et spontané de choix de saisie qui accompagne et renforce l'activité même de la mémoire du sujet. Les faits mémorisés ne sont pas choisis par un processus automatique et systématique, mais élaborés par le sujet en perpétuelle évolution.

Enfin nous avons insisté sur le fait qu'une exo-mémoire doit rester totalement privée et que seules les informations choisies par son propriétaire peuvent être communiquées à l'extérieur ainsi qu'il en est d'ailleurs de la mémoire biologique du sujet. Une exo-mémoire n'a aucune autre fonction sociale que d'apporter une aide à son propriétaire.

L'approche présentée ici s'apparente plus à un outil de travail sur soi [19], bloc-note ou hypomnema personnel. Certes s'il est vrai, comme l'affirme Michel Serres [20], que les nouveaux moyens technologiques engendrent des formes de néo-darwinisme¹³, ces moyens devraient également contribuer à favoriser un travail sur soi, toujours intime et indispensable, tout en bénéficiant des apports technologiques modernes.

Références

1. Delahaye, J.P. : Complexités. Aux limites des mathématiques et de l'informatique. Belin - pour la science (2006)
2. Fuller, M., Kelly, L., Jones, G.J.F. : Applying contextual memory cues for retrieval from personal information archives. In : PIM 2008 - Proceedings of Personal Information Management, Workshop at CHI. (2008)

13. Très simplement dit, l'idée est que les moyens de stockage modernes de l'information rendent inutile de concentrer l'effort cérébral humain sur la mémorisation pure, et permettent de libérer ainsi de nouvelles fonctionnalités.

3. Coulin, D. : Les traces. Editions Bernard Grasset, Paris (2004)
4. Bush, V. : As we May Think. The Atlantic Monthly (1945)
The electronic version was prepared by Denys Duchier, April 1994,
<http://ccat.sas.upenn.edu/~jod/texts/vannevar.bush.html>.
5. Chapouthier, G. : Biologie de la mémoire. Odile Jacob (2006)
6. Bell, G., ed. : The 3rd ACM Workshop on Capture, Archival and Retrieval of Personal Experiences, Santa Barbara, California, USA. In Bell, G., ed. : CARPE, ACM (2006) First workshop :
<http://research.microsoft.com/CARPE2004/schedule.htm>.
7. Vemuri, S., Bender, W. : Next-generation personal memory aids. BT Technology Journal **22** (2004)
8. Deransart, P. : Conception de Trace et Applications (vers une méta-théorie des traces). Technical report, Inria Paris-Rocquencourt (2009) Working document
<http://hal.inria.fr/>.
9. Langevine, L., Deransart, P., Ducassé, M. : A generic trace schema for the portability of cp(fd) debugging tools. In Apt, K., Fages, F., Rossi, F., Szeredi, P., Vancza, J., eds. : Recent Advances in Constraints. Number 3010 in LNAI. Springer Verlag (2004)
10. Deransart, P., Ducassé, M., Ferrand, G. : Observational semantics of the resolution box model. In Vanhoof, W., Hill, P., eds. : Proceedings of the 17th Workshop on Logic-based Methods in Programming Environments (WLPE'07), a post-conference workshop of ICLP'07, Porto, Portugal (2007) **à paraître dans le Computing Research Repository (CoRR)**.
11. Marty, J.C., Mille, A. : Analyse de traces et personnalisation des environnements informatiques pour l'apprentissage humain. Hermès, Lavoisier (2009)
12. Mille, A. : From case-based reasoning to trace-based reasoning. Annual Reviews in Control **2** (2006) 223–232
13. Edelman, G.M., Tononi, G. : A universe of Consciousness. How Matter becomes Imagination. Basic Books (2000) French translation : "Comment la matière devient conscience", Odile Jacob, 2000.
14. Matthiew, L.L., Dey, A.K. : Lifelogging memory appliance for people with episodic memory impairment. In : Proceedings of UbiComp'08, Seoul, Korea, ACM (2008)
15. Vemuri, S., Schmandt, C., Bender, W. : iRemember : a Personal, Long-term Memory Prosthesis. In ACM, ed. : Proceedings of the second ACM Workshop on Continuous Archival and Retrieval of Personal Experiences CARPE'06, Santa Barbara, California, USA (2006) 65–74
16. Giurca, A., Gasevic, D., Taveter, K. : Handbook of Research on Emerging Rule-Based Languages and Technologies, Open Solutions and Approaches. Information Science Reference (2009)
17. Collins, A.M., Quillian, M.R. : Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior **8** (1969) 240–248
18. Kiss, A., Quinqueton, J. : UNISCRIP T : a Model for Persistent and Incremental Knowledge Storage. In : The First ACM Workshop on Continuous Archival and Retrieval of Personal Experiences, CARPE 2004, Columbia University New York, Microsoft (2004)

19. Foucault, M. : L'écriture de soi. Dits et écrits **4** (1984) 415–431 1980-1988, 912 pages, 140 x 225 mm. Collection Bibliothèque des Sciences humaines, ISBN 2070739899.
20. Serres, M. : Les nouvelles technologies, que nous apportent-elles ? Interstice (2006) Michel Serres's conference recorded at "Ecole Polytechnique", decembre 1rst, 2005.